



Strategy of tuning gene expression ratio in prokaryotic cell from perspective of noise and correlation

Rui Li ^a, Liufang Xu ^b, Hualin Shi ^{a,*}

^a State Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China

^b Department of Physics and Biophysics & Complex System Center, Jilin University, Changchun 130012, China



HIGHLIGHTS

- Strategy of tuning gene expression ratio from perspective of noise and correlation is discussed.
- Analytical results of genes expression noise and correlation for the two-state and the constitutive promoter are given.
- Noises and correlation of gene expressions in operon with the two-state promoter are higher than in operon with the constitutive promoter.
- Polar effect obviously lowers the correlation between genes in operons.
- Tuning translation rates is the optimal strategy to tune gene expression ratio in operon.

ARTICLE INFO

Article history:

Received 1 April 2014

Received in revised form

31 October 2014

Accepted 3 November 2014

Available online 11 November 2014

Keywords:

Polar effect

The two-state promoter

Transcription terminator

Gene expression correlation

Operon structure

ABSTRACT

Genes are organized into operons in prokaryote, and these genes in one operon generally have related functions. However, genes in the same operon are usually not equally expressed, and the ratio needs to be fine-tuned for specific functions. We examine the difference of gene expression noise and correlation when tuning the expression level at the transcriptional or translational level in a bicistronic operon driven by a constitutive or a two-state promoter. We get analytic results for the noise and correlation of gene expression levels, which is confirmed by our stochastic simulations. Both the noise and the correlation of gene expressions in an operon with a two-state promoter are higher than in an operon with a constitutive promoter. Premature termination of mRNA induced by transcription terminator in the intergenic region or changing translation rates can tune the protein ratio at the transcriptional level or at the translational level. We find that gene expression correlation between promoter-proximal and promoter-distal genes at the protein level decreases as termination increases. In contrast, changing translation rates in the normal range almost does not alter the correlation. This explains why the translation rate is a key factor of modulating gene expressions in an operon. Our results can be useful to understand the relationship between the operon structure and the biological function of a gene network, and also may help in synthetic biology design.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

An operon is a unit of genomic DNA containing a cluster of genes under the control of a single promoter (Jacob and Monod, 1961). Genes in an operon generally belong to the same function class (Salgado and Moreno-Hagelsieb, 2000; Dandekar et al., 1998), and they are transcribed together into a mRNA strand. Their expressions can be regulated coordinately, which is one of the reasons that genes are organized into operons (Price et al., 2005).

Gene expression is noisy, which can influence its function (Eldar and Elowitz, 2010; Hao et al., 2011a; Paulsson et al., 2000; Ray and Igoshin, 2012; Chalancon et al., 2012). Noise from the stochasticity of transcription and translation process is recognized as the intrinsic noise. Noise introduced by other factors such as different concentrations of RNA polymerase, ribosome or transcription factors is recognized as the extrinsic noise. The extrinsic noise is relatively slow in the same environment (Taniguchi et al., 2010; Rosenfeld et al., 2005). Intrinsic and extrinsic noises together determine the gene expression profile (Taniguchi et al., 2010; Rosenfeld et al., 2005; Paulsson, 2004).

Proteins work together to execute specific functions in the cell. Different components in the same pathway need to be coordinated

* Corresponding author. Tel.: +86 10 62582506; fax: +86 10 62562587.
E-mail address: shih@itp.ac.cn (H. Shi).

to get the robust result. For example, protein complex formation is most efficient when subunits of a protein complex are expressed with the needed ratio (Carmi et al., 2006, 2009; Swain, 2004). Expressions of enzymes in a metabolism pathway usually need to be correlated to avoid intermediate product cumulation (Ray and Igoshin, 2012) which may be toxic. Experiment shows that gene expressions in the same operon are more correlated together than those in the separated operons even with identical promoters, especially when gene expression levels are low (Tabor et al., 2008). The strengthened correlation can be explained by co-transcription (Sneppen et al., 2010; Swain, 2004; Ray and Igoshin, 2012; Iber, 2006) or in some cases also by co-translation (Swain, 2004; Iber, 2006). Statistics from *Escherichia coli* genomics indicate that for genes with products in linear metabolic pathways, with physical interactions and in covalent modification modules, they are more likely to be organized into operons (Ray and Igoshin, 2012). Noise of the genes expression from one operon can be further attenuated by different mechanisms, for example, transcriptional or translational feedback (Swain, 2004). Sometimes subunits of macromolecular complex are produced from different operons. Swain (2004) shows that both translational and transcriptional feedback can increase the correlation between these subunits, but translational feedback is more efficient.

However, even in the same operon, genes are not expressed at the equivalent level. The cell needs to fine-tune the gene expression ratio to fit their biological functions. The tuning can be executed at the transcriptional or the translational level. In this work, we try to study the difference of noise and correlation when the cell takes distinct tuning strategies.

Tuning transcription rate or stability of mRNA can change the mRNA level, which will be passed to the protein levels by translation. Lifetime of mRNAs in bacteria is relatively short. Most mRNAs will be degraded by various ribonucleases (Condon, 2007) and live only several minutes (Selinger et al., 2003; Bernstein et al., 2002; Kristoffersen et al., 2012). Different transcription rates can be introduced by the internal promoter or terminator. Intergenic Rho-dependent or Rho-independent (Lesnik et al., 2001) terminator can introduce the polar effect which means genes in the promoter-proximal region are transcribed more compared with the promoter-distal ones. Recently, transcriptome sequencing of *Mycoplasma pneumoniae* (Güell et al., 2009) and synthetic biology studies in *E. coli* (Nishizaki et al., 2007; Hiroe et al., 2012) indicate that this effect might be a global phenomenon even for an operon without an obvious termination signal in the intergenic region. As for mRNA degradation rate, although distinct regions in one operon may have different degradation rates (Selinger et al., 2003; Kristoffersen et al., 2012), for those genes with close function relationship, their transcripts seem to have the similar degradation rates (Bernstein et al., 2002; Kristoffersen et al., 2012; Quax et al., 2013). Previous works indicate that co-transcription of genes in one operon is the main source of correlation (Sneppen et al., 2010; Swain, 2004; Ray and Igoshin, 2012; Iber, 2006), so it can be expected that tuning gene expression ratio at transcriptional level would lower the correlation.

Although it cannot be ruled out that tuning translation degradation would be one of the mechanisms to adjust the protein expression level, it is well known that most proteins in the prokaryotic cell are not degraded to avoid wasting (Gur et al., 2011). It is more likely that cells tune a protein level by adjusting protein synthesis rate. Many features embedded in coding and non-coding regions of the mRNA sequence can influence the translation initiation and elongation rate. The Shine-Dalgarno sequence and its distance to translation initiation codon will affect the assembling of translation complex (Vellanoeweth and Rabinowitz, 1992; Chen et al., 1994). The structure of 5'-UTR will affect the accessibility of the ribosome binding site (Goodman

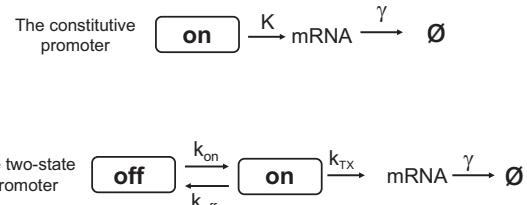
et al., 2013; Bentele et al., 2013; Hao et al., 2011b). Codon usage will influence translation elongation (Tuller et al., 2010; Cannarozzi et al., 2010; Klumpp et al., 2012). The internal Shine-Dalgarno like sequence may induce translation pausing (Li et al., 2012; Wen et al., 2008).

In this work, we use stochastic models to describe the transcription and translation processes of a bicistronic operon driven by a constitutive or a two-state promoter (Fig. 1A) respectively. We get analytical results of intrinsic noise and correlation of gene expressions. We utilize termination efficiency of terminator to characterize the stiffness of polar effect, i.e., the ratio of transcripts that are prematurely terminated. Stochastic simulations confirm our analytical results. Our results show that the correlation between proteins decreases monotonously as termination efficiency increases and the correlation almost does not vary as the translation rate changes in the normal range. It may help to understand gene organization and its evolution in prokaryotes.

2. The bicistronic operon with the constitutive promoter

In our analysis, we focus on the statistics when the system is in the steady state. We use $\langle q \rangle$ and δq^2 to denote the mean and variance of numbers of species q , the Fano factor $\nu = \delta q^2 / \langle q \rangle$ and coefficient of variance $\eta = \sqrt{\delta q^2 / \langle q \rangle}$ to characterize noise. The Fano factor can be a measurement for deviation from the Poisson behavior. It takes the value of 1 for a Poisson process. We use the Pearson correlation coefficient $\rho = \langle (q_1 - \langle q_1 \rangle)(q_2 - \langle q_2 \rangle) \rangle / \sqrt{\delta q_1^2 \delta q_2^2}$ to denote the correlation of the numbers of two species q_1 and q_2 .

A



B

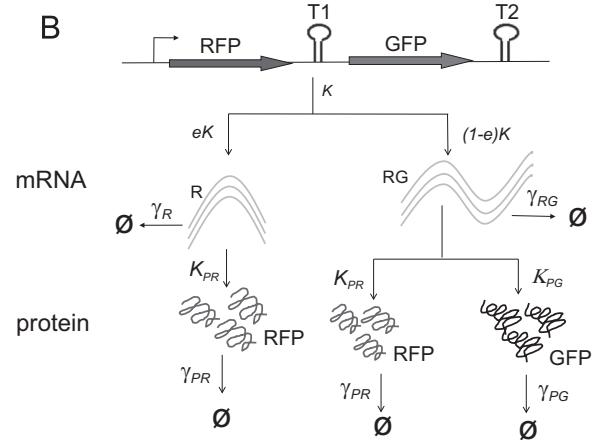
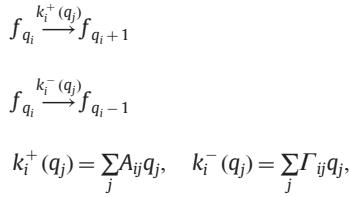


Fig. 1. (A) Two kinds of promoters used in our study. (B) Gene expression process of the bicistronic operon with the constitutive promoter. T1 represents any kind of polar effect mechanism: Rho-dependent or Rho-independent terminator in intergenic region of operon or other possible mechanisms (Güell et al., 2009). The termination efficiency of T1 is e , the transcription rate of promoter is K , the translation rates of RFP and GFP are K_{PR} and K_{PG} , the degradation rates of R, RG, RFP and GFP are respectively γ_R , γ_{RG} , γ_{PR} and γ_{PG} . For the two-state promoter, K should be replaced with k_{TX} , and transcription happens only when the promoter is in the on state.

2.1. The model of the bicistronic operon with the constitutive promoter

The state of a gene expression system can be specified by the numbers of mRNA molecules and proteins. For the stochasticity of gene expression, there are fluctuations in these numbers. We can describe this kind system by a joint distribution function $f(q_1, q_2, \dots, q_n; t)$, here q_i is the number of i -th molecule, which will be mRNA or protein. We simply write $f(q_1, q_2, \dots, q_n; t)$ as f_{q_i} , and i is ranging from 1 to n . For simplicity, we suppose that the birth and death rates of molecules in the system are linearly dependent on the numbers of different species. Then dynamic evolution of f_{q_i} is described in the scheme:



where $k_i^+(q_j)$ and $k_i^-(q_j)$ denote the birth and death rates of i -th molecule, matrices A and Γ provide the coefficients of their linear dependence on the molecular numbers.

We study the system of a bicistronic operon as described in Fig. 1B. We use *RFP* and *GFP* to denote genes in the operon with *RFP* near the promoter. These two fluorescent proteins can be employed in an experiment to check our theoretical results. T1 is Rho-dependent or independent terminator or any other possible mechanism that would attenuate downstream gene transcription. T2 is the terminator that defines the end of the operon. Termination efficiency of T1 is denoted as e , which also represents the stiffness of polar effect for the operon. For simplicity, we suppose that transcription and translation rates do not depend on each other. Although transcription and translation are coupled together in prokaryote, this assumption is still suitable for most of the cases.

We first study the bicistronic operon with the constitutive promoter. Transcription start from the constitutive promoter at the rate K as shown in Fig. 1B. Two kinds of mRNA are transcribed from this operon. One is full-length mRNA with both *RFP* and *GFP*, which is denoted as *RG*. The other one is attenuated mRNA only with *RFP*, which is denoted as *R*. The synthesis rate of *R* and *RG* are $K_R = eK$ and $K_{RG} = (1-e)K$. The translation rates of *RFP* and *GFP* are K_{PR} and K_{PG} , the degradation rates of *R*, *RG*, *RFP* and *GFP* are respectively γ_R , γ_{RG} , γ_{PR} and γ_{PG} . If we consider the growth dilution, the effective degradation rate $\gamma = \gamma_{degrad} + \mu$, where γ_{degrad} is the degradation rate and μ is the growth rate of the cell. In our system, $q_i = \{D, R, RG, RFP, GFP\}$, where D is the number of DNA, i.e., the copy number of the operon. Matrices A and Γ for this bicistronic operon with the constitutive promoter are:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ K_R & 0 & 0 & 0 & 0 \\ K_{RG} & 0 & 0 & 0 & 0 \\ 0 & K_{PR} & K_{PR} & 0 & 0 \\ 0 & 0 & K_{PG} & 0 & 0 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma_R & 0 & 0 & 0 \\ 0 & 0 & \gamma_{RG} & 0 & 0 \\ 0 & 0 & 0 & \gamma_{PR} & 0 \\ 0 & 0 & 0 & 0 & \gamma_{PG} \end{bmatrix}. \quad (1)$$

Transcription process of the bicistronic operon with the constitutive promoter includes a transcription initiation at rate K followed by a termination at the T1 terminator with a possibility of e (Fig. 1B). In our model, we used an equivalent description: two independent transcription processes of *R* and *RG* with the rates of $K_R = eK$ and $K_{RG} = (1-e)K$. It is easy to prove that these two descriptions give the same distributions of *R* and *RG*: two independent Poisson distributions.

2.1.1. Method to get the means and variances.

We use a similar approach as Thattai and van Oudenaarden (2001) in our theoretical derivations. The above scheme for our model can be described by the following Master Equation:

$$\dot{f}_{q_i} = (E_i^{-1} - 1) \left(\sum_j A_{ij} q_j \right) f_{q_i} + (E_i^{+1} - 1) \left(\sum_j \Gamma_{ij} q_j \right) f_{q_i}, \quad (2)$$

where E is the step operator defined as $E_i^k f(\dots, q_i, \dots) = f(\dots, q_i + k, \dots)$. This equation can be solved using generating function method. First we denote generating function $F(z_j, t) = \sum_{q_\nu=1,\dots,\infty} (\prod_{\nu=1,\dots,n} z_\nu^{q_\nu}) f_{q_\nu}$. In our system, Γ is diagonal: $\Gamma_{ij} = \delta_{ij} \Gamma_i$, and then \dot{F} obeys

$$\dot{F} = \sum_i (1 - z_i) \left(\Gamma_i F_i - \sum_j A_{ij} z_j F_j \right), \quad (3)$$

where $F_j = \partial z_j F$. In the steady state, $\dot{F} = 0$.

The mean and variance of species will be given by following properties of generating function: $F|_1 = 1; F_j|_1 = \langle q_j \rangle; F_{ii}|_1 = \langle q_i^2 \rangle - \langle q_i \rangle^2$; and if $i \neq j$, $F_{ij}|_1 = \langle q_i q_j \rangle$, where $F_{ij} = \partial z_i \partial z_j F$ and $|_1$ denotes evaluation at $z_j = 1$ for all j . Successive differentiations of Eq. (3) will generate linear equations of higher order moments. First we define the vector $J_i = F_i$, and the matrix $K_{ij} = F_{ij}$, where the indexes of K mean row and column and the indexes of F mean differentiation. Differentiating Eq. (3) up to the second moment, we obtain the following equations for the steady state:

$$(A - \Gamma)J = 0, \quad [(A - \Gamma)K + L] = -[(A - \Gamma)K + L]^T, \quad (4)$$

where $L_{ij} = A_{ij} J_j$ (not summation) (Thattai and van Oudenaarden, 2001).

For our bicistronic operon model, $J = \langle D \rangle, \langle R \rangle, \langle RG \rangle, \langle RFP \rangle, \langle GFP \rangle$. By solving Eq. (4), we got the mean (J) and variance(K) of gene expression, and the correlation of different genes in the steady state.

2.2. Results of the bicistronic operon with the constitutive promoter

2.2.1. Statistics of mRNAs

Both *R* and *RG* obey the Poisson distribution, which means their Fano factors are equal to 1. They are also independent with each other. The average numbers of *R* and *RG* are

$$\langle R \rangle = \frac{K_R}{Y_R} = \frac{eK}{\gamma_R}, \quad (5)$$

$$\langle RG \rangle = \frac{K_{RG}}{Y_{RG}} = \frac{(1-e)K}{\gamma_{RG}}. \quad (6)$$

Coefficient of variance of mRNAs are

$$\eta_R = \frac{\sqrt{\delta R^2}}{\langle R \rangle} = \sqrt{\frac{\gamma_R}{eK}} = \frac{1}{\sqrt{\langle R \rangle}}, \quad (7)$$

$$\eta_{RG} = \frac{\sqrt{\delta RG^2}}{\langle RG \rangle} = \sqrt{\frac{\gamma_{RG}}{(1-e)K}} = \frac{1}{\sqrt{\langle RG \rangle}}. \quad (8)$$

From these two equations, we can see that lower level of mRNA expression leads to larger fluctuation.

2.2.2. Statistics of proteins

During the short lifetime of the mRNA, proteins are synthesized by ribosomes with the mRNA, so the profile of translation of protein is bursty as shown in Fig. 2B. This is one of the sources of the intrinsic noise. The number of protein produced from a single mRNA obeys the geometric distribution (Berg, 1978). We use b to denote the burstiness of the translation process. It is the average number of protein translated by ribosomes during the lifetime of one mRNA, for example, $b_{PR-R} = K_{PR}/\gamma_R$ is the average number of

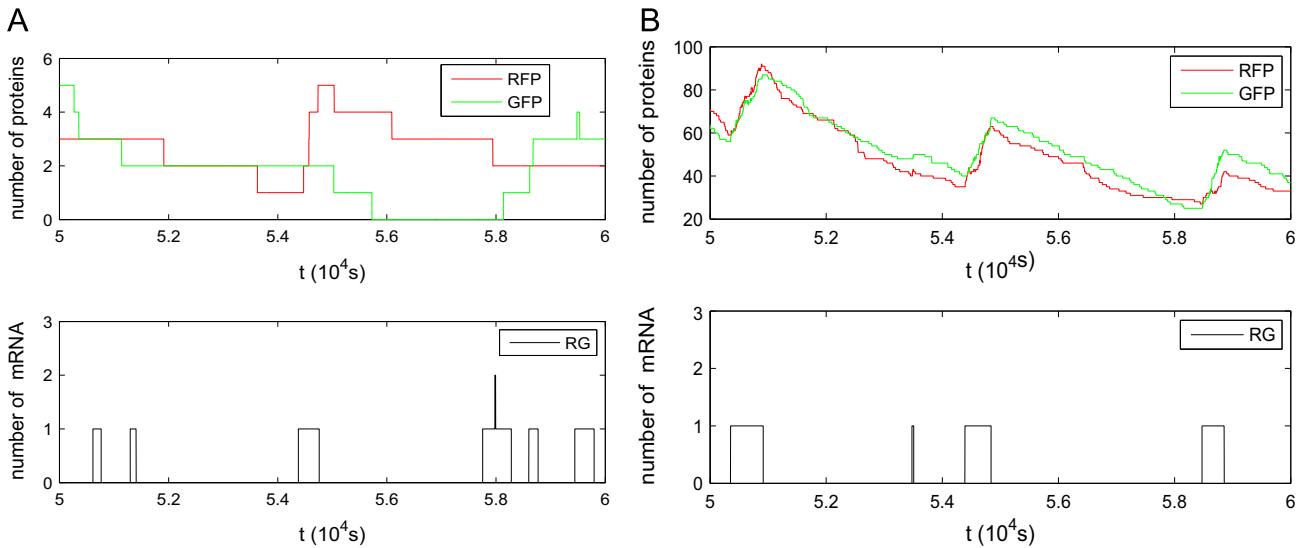


Fig. 2. Sampling of gene expression profiles from stochastic simulation for the operon with the constitutive promoter. Small translation burst sizes in (A) would lower the correlation. In the simulations $\gamma_{PR} = \gamma_{PG} = 0.00029 \text{ s}^{-1}$, $\gamma_R = \gamma_{RG} = 0.0033 \text{ s}^{-1}$, $K = 0.0005 \text{ s}^{-1}$, $e = 0$, other parameters in (A) $K_{PR} = K_{PG} = 0.0033 \text{ s}^{-1}$, the average translational burst size of RFP and GFP are $b_{PR} = b_{PG} = 1$; (B) $K_{PR} = K_{PG} = 0.066 \text{ s}^{-1}$, $b_{PR} = b_{PG} = 20$.

RFP protein translated in the lifetime of one RG mRNA. Similarly, we denote $b_{PR-RG} = K_{PR}/\gamma_{RG}$ and $b_{PG} = K_{PG}/\gamma_{RG}$. In the cell, protein lives much longer than mRNA, i.e., $\gamma_{PR}, \gamma_{PG} \ll \gamma_R, \gamma_{RG}$, so we made approximations under this assumption below. Average numbers of proteins are

$$\begin{aligned} \langle RFP \rangle &= \frac{K_{PR}}{\gamma_{PR}}(\langle R \rangle + \langle RG \rangle) = \frac{K_{PR}}{\gamma_{PR}} \left(\frac{K_R}{\gamma_R} + \frac{K_{RG}}{\gamma_{RG}} \right) \\ &= \frac{K}{\gamma_{PR}} [e b_{PR-R} + (1-e) b_{PR-RG}], \end{aligned} \quad (9)$$

$$\langle GFP \rangle = \frac{K_{PG}}{\gamma_{PG}} \langle RG \rangle = \frac{K_{PG} K_{RG}}{\gamma_{PG} \gamma_{RG}} = \frac{(1-e)K}{\gamma_{PG}} b_{PG}. \quad (10)$$

The ratio between them is

$$\frac{\langle RFP \rangle}{\langle GFP \rangle} = \frac{K_{PR}/K_{PG}}{\gamma_{PR}/\gamma_{PG}} \left[\frac{e/(1-e)}{\gamma_R/\gamma_{RG}} + 1 \right]. \quad (11)$$

We can see different strategies tuning protein expression ratio in one operon: changing termination efficiency e , changing the translation rate ratio K_{PR}/K_{PG} , changing the protein or mRNA stability ratio, i.e., γ_{PR}/γ_{PG} , or γ_R/γ_{RG} . Although tuning protein ratio at protein degradation rates cannot be ruled out, it seems that the prokaryotic cell avoids to degrade protein to avoid wasting (Gur et al., 2011). For functional related proteins in the same operon, especially for those without transcriptional regulatory function, mRNAs tend to have the similar lifetimes (Bernstein et al., 2002; Kristoffersen et al., 2012; Quax et al., 2013). It seems that adjusting translation rates or termination efficiency is widely used mechanism in the prokaryotic cell to tune the gene expression ratio (Quax et al., 2013; Cho et al., 2009; Güell et al., 2009).

The Fano factor of GFP is

$$\frac{\delta GFP^2}{\langle GFP \rangle} = \frac{K_{PG}}{\gamma_{PG} + \gamma_{RG}} + 1 \approx \frac{K_{PG}}{\gamma_{RG}} + 1 = b_{PG} + 1, \quad (12)$$

which is the same as in Thattai and van Oudenaarden (2001). The distribution for the molecule numbers of protein is broader than the Poisson distribution.

Coefficient of Variance of GFP protein is

$$\eta_{GFP} = \sqrt{\frac{1}{\langle GFP \rangle} + \frac{1}{1 + \frac{\gamma_{RG}}{\gamma_{PG}} \langle RG \rangle}}$$

$$\begin{aligned} &\approx \sqrt{\frac{1}{\langle GFP \rangle} + \frac{\gamma_{PG}}{(1-e)K}} \\ &= \sqrt{\frac{\gamma_{PG}}{(1-e)K} \left(\frac{1}{b_{PG}} + 1 \right)}. \end{aligned} \quad (13)$$

If we fix other parameters, the intrinsic noise of GFP is decreasing as the burst size b_{PG} increasing. When b_{PG} is small, burst of translation is a critical source of intrinsic noise, when b_{PG} is big, the intrinsic noise is mainly determined by the transcription process, i.e., by K and e , considering γ_{PG} is primarily determined by growth dilution. Increasing promoter strength (K) or decreasing premature termination (e) would repress mRNA expression noise as shown in Eq. (8), and then lower noise would be passed to the protein level. Our result agrees with Swain et al. (2002).

When $e \neq 0$ and $e \neq 1$, the Fano factor of protein RFP is

$$\begin{aligned} \frac{\delta RFP^2}{\langle RFP \rangle} &= \frac{K_{PR}}{(\gamma_{RG} + \gamma_{PR})} \left[\frac{\langle RG \rangle}{\langle RG \rangle + \langle R \rangle} \right] + \frac{K_{PR}}{(\gamma_R + \gamma_{PR})} \left[\frac{\langle R \rangle}{\langle RG \rangle + \langle R \rangle} \right] + 1 \\ &\approx b_{PR-RG} \left(\frac{1}{1 + \frac{e}{1-e} \frac{\gamma_{RG}}{\gamma_R}} \right) + b_{PR-R} \left(\frac{1}{1 + \frac{1-e}{e} \frac{\gamma_R}{\gamma_{RG}}} \right) + 1. \end{aligned} \quad (14)$$

Here we can see that the noise of RFP is derived from the Poisson processes of the translations from the two kinds of mRNAs. The ratio depends on their ratio in the total mRNA. When $e \rightarrow 1$, then

$$\frac{\delta RFP^2}{\langle RFP \rangle} \rightarrow \frac{K_{PR}}{(\gamma_R + \gamma_{PR})} + 1,$$

and when $e \rightarrow 0$, then

$$\frac{\delta RFP^2}{\langle RFP \rangle} \rightarrow \frac{K_{PR}}{(\gamma_{RG} + \gamma_{PR})} + 1.$$

Under these two conditions, RFP protein is produced from only R or RG , so RFP has the similar distribution as GFP (Eq. (12)).

When $\gamma_{RG} = \gamma_R$, $b_{PR-RG} = b_{PR-R} = b_{PR}$. And then the Fano factor of RFP can be simplified into

$$\frac{\delta RFP^2}{\langle RFP \rangle} = (1-e)b_{PR-RG} + eb_{PR-R} + 1 = b_{PR} + 1, \quad (15)$$

the Fano factor only depends on the translational burst size of RFP, which is in agreement with the well-known result for single gene expression (Thattai and van Oudenaarden, 2001).

Coefficient of variance of RFP protein is

$$\eta_{RFP} = \sqrt{\frac{1}{\langle RFP \rangle} + \left[\frac{1}{\left(1 + \frac{\gamma_{RG}}{\gamma_{PR}}\right)} \left(\frac{1}{1 + \frac{e}{1-e} \frac{\gamma_{RG}}{\gamma_R}} \right) + \frac{1}{\left(1 + \frac{\gamma_R}{\gamma_{PR}}\right)} \left(\frac{1}{1 + \frac{1-e}{e} \frac{\gamma_R}{\gamma_{RG}}} \right) \right] \frac{1}{\langle RG \rangle + \langle R \rangle}} \quad (16)$$

If $\gamma_R = \gamma_{RG}$, then

$$\begin{aligned} \eta_{RFP} &= \sqrt{\frac{1}{\langle RFP \rangle} + \frac{1}{\left(1 + \frac{\gamma_{RG}}{\gamma_{PR}}\right)} \frac{1}{\langle RG \rangle + \langle R \rangle}} \\ &\approx \sqrt{\frac{1}{\langle RFP \rangle} + \frac{\gamma_{PR}}{K}} \\ &= \sqrt{\frac{\gamma_{PR}}{K} \left(\frac{1}{b_{PR}} + 1 \right)}, \end{aligned} \quad (17)$$

where $b_{PR} = b_{PR-R} = b_{PR-RG}$. RFP can be translated equally from both R and RG , and these two kinds mRNAs have the same degradation rates as assumed. So termination efficiency (e) would not influence the noise of RFP. From Eqs. (13) and (17), clearly noise of RFP comes from: (1) noise inherited from the transcription process which is measured by transcription rate $1/K$; (2) noise from burstiness of translation measured by burst size b_{PR} . When transcription rate K or burst size b_{PR} is small, the fluctuation of RFP expression would be significant.

2.2.3. The correlation between RFP and GFP mRNAs

In order to see how much correlation at the mRNA level would be passed to the protein level, we first analyze the correlation between numbers of mRNAs that contains RFP or GFP.

$$\begin{aligned} \rho_{mR-mG} &= \frac{\langle (RG+R) - \langle RG+R \rangle \rangle \langle (RG-R) \rangle}{\sqrt{\delta(RG+R)^2} \sqrt{\delta RG^2}} \\ &= \frac{1}{\sqrt{1 + \frac{\delta R^2}{\delta RG^2}}} \\ &= \sqrt{\frac{\langle RG \rangle}{\langle RG \rangle + \langle R \rangle}} \\ &= 1 / \sqrt{1 + \frac{e}{1-e} \frac{\gamma_{RG}}{\gamma_R}}, \end{aligned} \quad (18)$$

which is square root of the ratio of RG mRNA in the total mRNA. This equation shows that the correlation between mRNAs that contains RFP and GFP is from co-transcription of RFP and GFP. Longer lifetime of RG relative to R will lead to stronger correlation.

If $\gamma_{RG} = \gamma_R$, then

$$\rho_{mR-mG} = \sqrt{1-e}, \quad (19)$$

the correlation between RFP and GFP at the mRNA level is only dependent on the termination efficiency e of the terminator.

2.2.4. The correlation between mRNA and protein

We then calculate the correlation between mRNA and protein expressions. The Pearson correlation coefficient between RG and GFP is

$$\rho_{mG-GFP} = \sqrt{\frac{1}{\left(1 + \frac{1}{K_{PG}/\gamma_{RG}} + \frac{1}{K_{PG}/\gamma_{PG}}\right) \left(1 + \frac{\gamma_{RG}}{\gamma_{PG}}\right)}}, \quad (20)$$

where mG is RG , which is the only mRNA coding GFP protein. Proteins live much longer than mRNAs, which means

$$\frac{1}{K_{PG}/\gamma_{RG}} \gg \frac{1}{K_{PG}/\gamma_{PR}} \quad \text{and} \quad \frac{\gamma_{PG}}{\gamma_{RG}} \approx 0,$$

then

$$\rho_{RG-GFP} \approx \sqrt{\frac{\gamma_{PG}/\gamma_{RG}}{1 + \frac{1}{K_{PG}/\gamma_{RG}}}} \approx 0. \quad (21)$$

When $\gamma_{RG} = \gamma_R$ and $\gamma_{PR} = \gamma_{PG}$, we got the correlation between mRNAs and RFP protein:

$$\rho_{mR-RFP} = \sqrt{\frac{1}{\left(1 + \frac{1}{K_{PR}/\gamma_R} + \frac{1}{K_{PR}/\gamma_{PR}}\right) \left(1 + \frac{\gamma_R}{\gamma_{PR}}\right)}} \approx \sqrt{\frac{\gamma_{PR}/\gamma_R}{1 + \frac{1}{K_{PR}/\gamma_R}}} \approx 0, \quad (22)$$

where mR indicates the total number of transcripts containing RFP coding sequence, which include full-length transcripts RG and short ones R . From Eqs. (21) and (22), we see that there is almost no correlation between mRNA and protein due to $\gamma_{protein} \ll \gamma_{mRNA}$. Actually, experiments show that protein levels do not or just weakly correlate with mRNA levels (Taniguchi et al., 2010; Golding et al., 2005). Due to the much longer lifetime of protein, the fluctuation at the mRNA level is smoothed out at the protein level, which leads to the weak correlation between mRNA and protein.

2.2.5. The correlation between RFP and GFP proteins

Next we calculate gene expression correlation between two proteins, which has the biology significance for large protein complex formation, and for reaction systems sensitive to enzyme correlation (Ray and Igoshin, 2012). The Pearson correlation coefficient of the numbers of RFP and GFP protein is

$$\rho_{RFP-GFP} = \frac{\langle RFP \cdot GFP \rangle - \langle RFP \rangle \langle GFP \rangle}{\sqrt{\delta RFP^2} \sqrt{\delta GFP^2}}. \quad (23)$$

The general analytic formula is a little complex, and we do not show it here. In order to get intuitive understanding of the dependence of $\rho_{RFP-GFP}$ on different parameters, we make simplifications below.

If $\gamma_{RG} = \gamma_R$, $\gamma_{PR} = \gamma_{PG}$, then the correlation coefficient of RFP and GFP can be simplified into

$$\begin{aligned} \rho_{RFP-GFP} &= \frac{\sqrt{1-e}}{\sqrt{\left(1 + \frac{1}{K_{PR}/\gamma_R} + \frac{1}{K_{PR}/\gamma_{PR}}\right) \left(1 + \frac{1}{K_{PG}/\gamma_{RG}} + \frac{1}{K_{PG}/\gamma_{PG}}\right)}}} \\ &\approx \frac{\sqrt{1-e}}{\sqrt{\left(1 + \frac{1}{K_{PR}/\gamma_R}\right) \left(1 + \frac{1}{K_{PG}/\gamma_{RG}}\right)}} \\ &= \frac{\sqrt{1-e}}{\sqrt{(1+1/b_{PR})(1+1/b_{PG})}}, \end{aligned} \quad (24)$$

where $b_{PR} = b_{PR-R} = b_{PR-RG}$, for $\gamma_{RG} = \gamma_R$. Changing promoter strength (K) would not influence the correlation, although it can influence noise of mRNA and protein. When $e=0$, our result is the same as Ray and Igoshin (2012). In Eq. (24), the numerator $\sqrt{1-e}$ represents protein correlation inherited from the mRNA level (see Eq. (19)) and the denominator shows the influence from the bursty translation process. If b_{RFP} or b_{GFP} is small, then the number of RFP or GFP produced in the lifetime of one RG has large fluctuation (Eq. (17)). But translation events of RFP and GFP are independent and their fluctuations are unrelated, which means correlation between RFP and GFP would be low. This has been shown by the stochastic simulations in Fig. 2A (simulation method see Appendix B). However, when both of b_{RFP} and b_{GFP} are much larger than 1, fluctuations of RFP or GFP are both small (Eqs. (13) and (17)). So

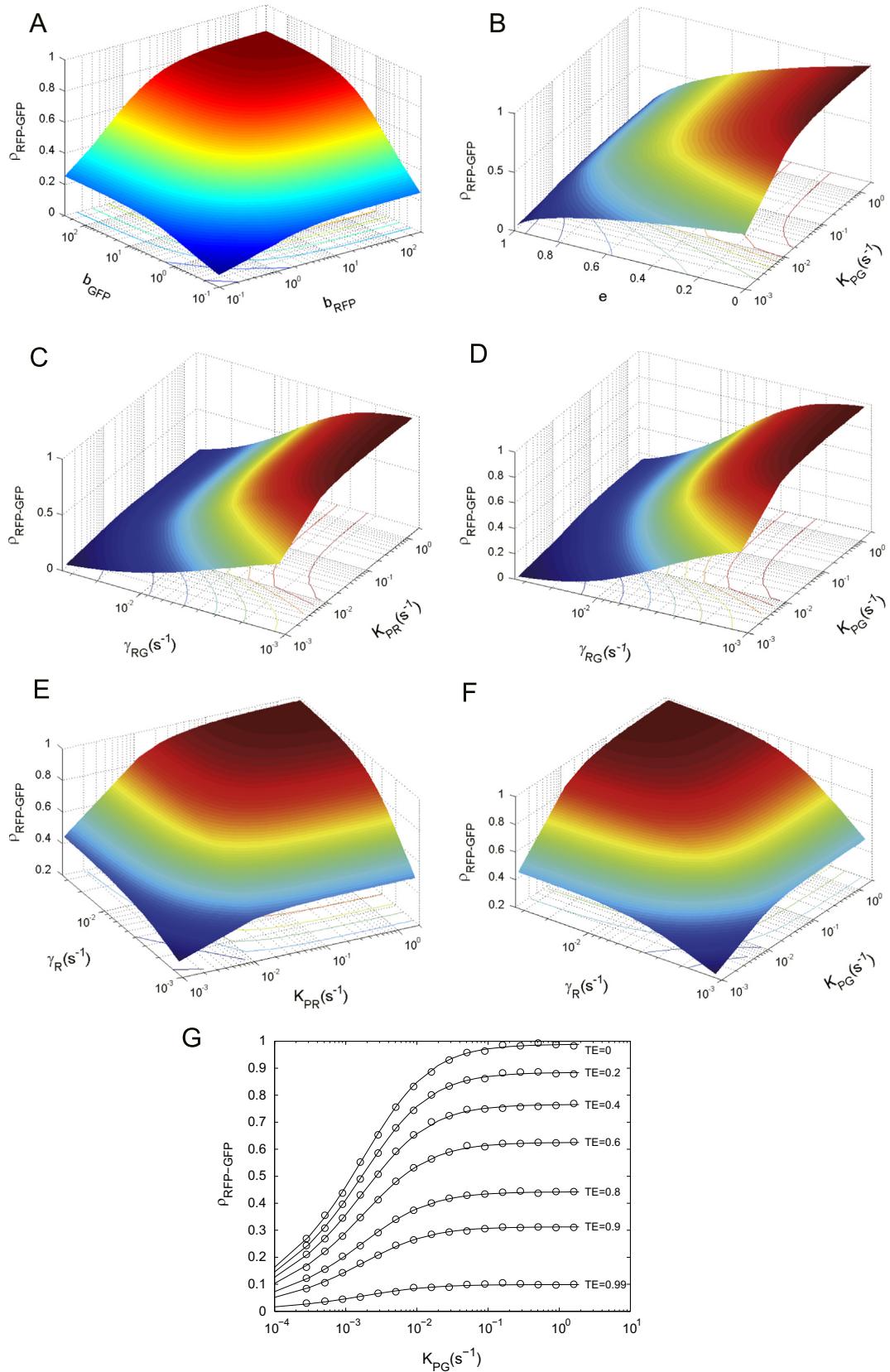


Fig. 3. Dependence of the correlation between RFP and GFP on terminator efficiency e , translation burst size b , translation rates of proteins and the degradation rates of mRNAs in bicistronic operon with the constitutive promoter. (A–F) all data are obtained from the analytical results, lines in surface are contours of $\rho_{RFP-GFP}$. (G) Cycles are got from stochastic simulations and lines are got from analytical results. In all figures $\gamma_{PR} = \gamma_{PG} = 0.00029 s^{-1}$, other parameters in each subplot are specified as follows: in (A) $e = 0.3, \gamma_R = \gamma_{RG} = 0.0033 s^{-1}$; (B) $\gamma_R = \gamma_{RG} = 0.0033 s^{-1}, K_{PR} = 0.16 s^{-1}$; (C) $e = 0.3, \gamma_R = 0.0033 s^{-1}, K_{PG} = 0.16 s^{-1}$; (D) $e = 0.3, \gamma_R = 0.0033 s^{-1}, K_{PR} = 0.16 s^{-1}$; (E) $e = 0.3, \gamma_{RG} = 0.0033 s^{-1}, K_{PG} = 0.16 s^{-1}$; (F) $e = 0.3, \gamma_{RG} = 0.0033 s^{-1}, K_{PR} = 0.16 s^{-1}$; (G) $K = 0.3 s^{-1}, K_{PR} = 0.16 s^{-1}, \gamma_{RG} = \gamma_R = 0.0033 s^{-1}$.

bursty translations have little effect on the correlation of two proteins, which has been shown by the simulation results in Fig. 2B. From Eq. (24) we can see that when both b_{PR} and b_{PG} are much bigger than 1 (Fig. 3A), $\rho_{RFP-GFP} \approx \sqrt{1-e}$, which means the correlation is mainly determined by the polar effect at the transcriptional level.

We plot the relationship between $\rho_{RFP-GFP}$ and other parameters in Fig. 3 using the formula of $\rho_{RFP-GFP}$ (Eq. (23)) without any simplification except supposing $\gamma_{PR} = \gamma_{PG}$. In Fig. 3G, we can see stochastic simulations agree with our analytical results very well. In Fig. 3B and G we can see that $\rho_{RFP-GFP}$ decreases as e increases, i.e., transcription termination at the internal terminator directly lowers the protein correlation. This result can explain why Monte Carlo simulation of Liang et al. (2013) only showed a small effect of intergenic terminator on the correlation. They take $e=0.5$, $\rho_{RFP-GFP} \approx \sqrt{1-e} \approx 0.7$ and the correlation is lowered just about 30%. Taking bigger e , for example, $e=0.9$, $\rho_{RFP-GFP} \approx \sqrt{1-e} \approx 0.32$, then the polar effect will lower the correlation about 70%, and the influence of polar effect would be obvious. In prokaryotic cells, mRNAs live about several minutes on average (Selinger et al., 2003; Bernstein et al., 2002; Kristoffersen et al., 2012). In the range of mRNA degradation rate (half-life of mRNA 0.5–15 min), adjusting the translation rate almost does not alter the protein correlation except translation rate is very small (Fig. 3B–G). This means adjusting protein expressions at the translation level will be the optimal strategy to tune the gene expression ratio in the same operon. If we take the half-life of mRNA 5 min, we can tune the translation rate between 0.01 and 1 s^{-1} without obviously changing the correlation of the proteins (Fig. 3G). Actually, comparative genomic study of Quax et al. (2013) shows that different translation rate is a key determinant of the gene expression ratio in an operon. Only those RFPs translated from RG contribute to the correlation with GFP. Decreasing the stability of mRNA R (increasing γ_R) or increasing the stability of mRNA RG (decreasing γ_{RG}) will increase the ratio of RFP correlated with GFP protein, which leads to a higher correlation (Fig. 3C–F). At the same time, increasing the stability of mRNA RG (decreasing γ_{RG}) will increase the burst size of correlated RFP and GFP translation from RG, which also leads to a higher correlation (Fig. 3C and D).

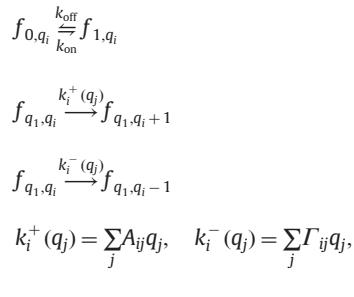
3. The bicistronic operon with the two-state promoter

Some single-cell experiments show that mRNA synthesis is also bursty for some promoters which can be described by the two-state model (Golding et al., 2005; So et al., 2011; Chong et al., 2014). As shown in Fig. 1A, these promoters stochastically transit between ‘on’ and ‘off’ states. The mRNA can be synthesized with the transcription rate k_{TX} only when the promoter is in the on state. The transition rate from the off state to the on state is k_{on} , which determines the average duration time of off state ($1/k_{on}$). The transition rate from the on state to the off state is k_{off} , which means the average duration of transcription pulse is $1/k_{off}$. The average transcription burst size, i.e., the average number of mRNA synthesized in one period of on state is k_{TX}/k_{off} . We study the gene expression noise in the bicistronic operon driven by the two-state promoter (Fig. 1). By solving the Master Equation, we get analytical results for the noises and the correlation at the mRNA level and the protein level. These analytical results are confirmed by stochastic simulation (procedure in Appendix B).

3.1. The model of the bicistronic operon with the two-state promoter

As in the system with the constitutive promoter, we use f_{q_i} ($i=1\dots n$) to describe the system. Here the state of the promoter is indicated by q_1 , $q_1=1$ for the on state and $q_1=0$ for the off

state. The molecular numbers of other species are indicated by q_i ($i=2\dots n$), and $q_i=0, 1\dots \infty$. We rewrite f_{q_i} as f_{q_1, q_i} ($i=2\dots n$), the dynamic evolution of f_{q_1, q_i} is described in the scheme:



where $k_i^+(q_j)$ and $k_i^-(q_j)$ denote the birth and death rates of i -th molecule, matrices A and Γ provide the coefficients of their linear dependence on the molecular numbers. In our systems, $q_i = \{D, R, RG, RFP, GFP\}$, where D is the number of active promoter, i.e., the number of the operon that can be transcribed. Matrices A and Γ are same as those we used for the operon with the constitutive promoter (Eq. (1)).

3.1.1. Method to get the means and variances

The above scheme for the two-state promoter can be described by the following Master Equation:

$$\begin{aligned} \dot{f}_{q_1, q_i} &= (E_i^{-1} - 1) \left(\sum_j A_{ij} q_j \right) f_{q_1, q_i} + (E_i^{+1} - 1) \left(\sum_j \Gamma_{ij} q_j \right) f_{q_1, q_i} \\ &\quad + (-1)^{q_1} (k_{off} f_{1, q_i} - k_{on} f_{0, q_i}). \end{aligned} \quad (25)$$

For the promoter in the on or the off state we denote the generating functions:

$$F_{on}(z_j, t) = \sum_{q_\nu=0\dots\infty} \left(\prod_{\nu=2\dots n} z_\nu^{q_\nu} \right) f_{1, q_\nu}, \quad (26)$$

$$F_{off}(z_j, t) = \sum_{q_\nu=0\dots\infty} \left(\prod_{\nu=2\dots n} z_\nu^{q_\nu} \right) f_{0, q_\nu}, \quad (27)$$

and for the whole system the generating function is $F(z_j, t) = z_1 F_{on}(z_j, t) + F_{off}(z_j, t)$. In our system, Γ is diagonal: $\Gamma_{ij} = \delta_{ij} \Gamma_i$, then F obeys

$$\begin{aligned} \dot{F} &= \sum_{i=2\dots n} (1-z_i) \left(\Gamma_i F_i - \sum_{j=1\dots n} A_{ij} z_j F_j \right) + z_1 (k_{on} F_{off} - k_{off} F_{on}) \\ &\quad + (k_{off} F_{on} - k_{on} F_{off}), \end{aligned} \quad (28)$$

where $F_j = \partial_{z_j} F$. In the steady state, $\dot{F} = 0$.

The mean and variance of species will be given by following properties of generating function: $F_j|_1 = \langle q_j \rangle$; $F_{ii} = \langle q_i^2 \rangle - \langle q_i \rangle^2$; and if $i \neq j$, $F_{ij}|_1 = F_{ji}|_1 = \langle q_i q_j \rangle$, where $F_{ij} = \partial_{z_i} \partial_{z_j} F$ and $|_1$ denotes evaluation at $z_j = 1$ for all j . Specially, $F_{11}|_1 = 0$; and if $i \neq 1$, then $F_{on}|_1 + F_{off}|_1 = \langle q_i \rangle$, $F_{on}|_1 = F_{11}|_1$. Successive differentiations of Eq. (28) will generate linear equations of higher order moments.

We denote $f_{on} = F_{on}|_1 = \sum_{q_\nu=0\dots\infty} f_{1, q_\nu}$ which is the possibility that promoter stays in the on state, and similarly $f_{off} = F_{off}|_1 = \sum_{q_\nu=0\dots\infty} f_{0, q_\nu}$ is the possibility of the off state. From the normalization of the possibility, we know that:

$$f_{on} + f_{off} = 1. \quad (29)$$

First we define the vector $J_i = F_i$, and the matrix $K_{ij} = F_{ij}$, where the indexes of K mean row and column and the indexes of F mean differentiation. Differentiating Eq. (28) up to the second moment, we obtain the following equations for the steady state:

$$k_{on} f_{off} - k_{off} f_{on} = 0, \quad (30)$$

$$\Gamma_i K_{i1} - \sum_{j=1 \dots n} A_{ij} K_{j1} - A_{i1} J_1 + k_{off} K_{1i} - k_{on} (J_i - K_{1i}) = 0 \quad (i = 2 \dots n), \quad (31)$$

$$(A - \Gamma) J = 0, \quad [(A - \Gamma) K + L]_{ij} = -[(A - \Gamma) K + L]_{ji} \quad (i, j = 2 \dots n), \quad (32)$$

where $L_{ij} = A_{ij} J_j$ (not summation).

3.2. Results of the bicistronic operon with the two-state promoter

By solving Eqs. (29)–(32) we get the noise and the correlation between molecules for the bicistronic operon with the two-state promoter. The results are shown below.

3.2.1. Statistics of mRNAs

The average number of mRNAs are

$$\langle R \rangle = \frac{ef_{on} k_{TX}}{\gamma_R}, \quad (33)$$

$$\langle RG \rangle = \frac{(1-e)f_{on} k_{TX}}{\gamma_{RG}}, \quad (34)$$

where $f_{on} = k_{on}/(k_{on} + k_{off})$ is the fraction of time or the possibility that the promoter is in the on state. If we denote $K = f_{on} k_{TX}$ as the effective transcription rate of the two-state promoter, then the average number of R and RG take the same formula as Eqs. (5) and (6) for the constitutive promoter.

The Fano factors of mRNAs are

$$\frac{\delta R^2}{\langle R \rangle} = 1 + eA_R, \quad (35)$$

$$\frac{\delta RG^2}{\langle RG \rangle} = 1 + (1-e)A_{RG}, \quad (36)$$

where

$$A_R = \frac{k_{off}}{k_{on} + k_{off}} \frac{k_{TX}}{(\gamma_R + k_{on} + k_{off})} \quad \text{and} \quad A_{RG} = \frac{k_{off}}{k_{on} + k_{off}} \frac{k_{TX}}{(\gamma_{RG} + k_{on} + k_{off})},$$

which characterize the influence of the two-state promoter. Our results of the Fano factors for the mRNAs are the same as previous results (Peccoud and Ycart, 1995). The Fano factors of R and RG are bigger than 1, which means the distribution of R and RG is not Poisson. Experimental measurements of mRNA show that the Fano factors of some mRNAs are greater than 1 (Golding et al., 2005). This implies that the two-state promoter model may be a more realistic description of these promoters (Golding et al., 2005; So et al., 2011; Chong et al., 2014).

Coefficients of Variance of mRNAs are

$$\eta_R = \frac{\sqrt{\delta R^2}}{\langle R \rangle} = \sqrt{\frac{1}{\langle R \rangle} (1 + eA_R)}, \quad (37)$$

$$\eta_{RG} = \frac{\sqrt{\delta RG^2}}{\langle RG \rangle} = \sqrt{\frac{1}{\langle RG \rangle} [1 + (1-e)A_{RG}]}. \quad (38)$$

The noise of mRNA depends on the mean expression level: the lower the expression level, the higher the noise. This behavior is similar to the mRNA transcribed from the constitutive promoter (Eqs. (7) and (8)). However if we keep the same average expression level of mRNA, the operon with the two-state promoter has higher noise of mRNA compared with the constitutive promoter. If we adjust the parameters of the two-state promoter and let $A_R, A_{RG} \rightarrow 0$, then the Fano factors and coefficients of variance of the mRNAs will be the same as those for the operon with the constitutive promoter (Eqs. (5)–(8)).

3.2.2. Statistics of proteins

The average expression levels of proteins are

$$\langle RFP \rangle = \frac{K_{PR}}{\gamma_{PR}} (\langle R \rangle + \langle RG \rangle) = \frac{f_{on} k_{TX}}{\gamma_{PR}} [eb_{PR-R} + (1-e)b_{PR-RG}], \quad (39)$$

$$\langle GFP \rangle = \frac{K_{PG}}{\gamma_{PG}} \langle RG \rangle = \frac{(1-e)f_{on} k_{TX}}{\gamma_{PG}} b_{PG}, \quad (40)$$

where $b_{PR-RG} = k_{PR}/\gamma_{RG}$, $b_{PR-R} = k_{PR}/\gamma_R$ and $b_{PG-RG} = k_{PG}/\gamma_{RG}$ are the translational burst sizes of RFP and GFP. If we use the effective transcription rate $K = f_{on} k_{TX}$ in Eqs. (39) and (40), the average expressions of the RFP and GFP for the two-state promoters take the same formula as those for the constitutive promoter (Eqs. (9) and (10)).

When $\gamma_{RG} = \gamma_R$, then $b_{PR-RG} = b_{PR-R} = b_{PR}$ and the Fano factors of GFP and RFP are

$$\frac{\delta GFP^2}{\langle GFP \rangle} = \frac{K_{PG}}{\gamma_{PG} + \gamma_{RG}} [1 + (1-e)B_G] + 1 \approx b_{PG} [1 + (1-e)B_G] + 1, \quad (41)$$

$$\frac{\delta RFP^2}{\langle RFP \rangle} = \frac{K_{PR}}{\gamma_{PR} + \gamma_{RG}} (1 + B_R) + 1 \approx b_{PR} (1 + B_R) + 1, \quad (42)$$

where

$$B_G = \frac{k_{off}}{k_{on} + k_{off}} \frac{k_{TX} (\gamma_{RG} + \gamma_{PR} + k_{on} + k_{off})}{(\gamma_{RG} + k_{on} + k_{off})(\gamma_{PR} + k_{on} + k_{off})}$$

and

$$B_R = \frac{k_{off}}{k_{on} + k_{off}} \frac{k_{TX} (\gamma_{RG} + \gamma_{PR} + k_{on} + k_{off})}{(\gamma_{RG} + k_{on} + k_{off})(\gamma_{PR} + k_{on} + k_{off})}.$$

We can see that the bigger the translational burst size, the bigger the Fano factors. The Fano factors of GFP and RFP (Eqs. (41) and (42)) for the two-state promoter are bigger than those for the constitutive promoter (Eqs. (12) and (14)), and the influence of the two-state promoter is characterized by B_G and B_R .

When $\gamma_{RG} = \gamma_R$, the coefficients of variance of GFP and RFP are

$$\begin{aligned} \eta_{GFP} &= \sqrt{\frac{1}{\langle GFP \rangle} \left(1 + \frac{k_{PG}}{\gamma_{RG} + \gamma_{PG}} \right) + \frac{1}{1 + \frac{\gamma_{RG}}{\gamma_{PG}} \langle RG \rangle} (1-e)B_G} \\ &\approx \sqrt{\frac{1}{\langle GFP \rangle} (1 + b_{PG}) + \frac{\gamma_{PG}}{(1-e)f_{on} k_{TX}} (1-e)B_G} \\ &= \sqrt{\frac{\gamma_{PG}}{(1-e)f_{on} k_{TX}} \left[\frac{1}{b_{PG}} + 1 + (1-e)B_G \right]}, \end{aligned} \quad (43)$$

$$\eta_{RFP} \approx \sqrt{\frac{\gamma_{PR}}{f_{on} k_{TX}} \left[\frac{1}{b_{PR}} + 1 + B_R \right]}. \quad (44)$$

We can see that noises of RFP and GFP are bigger than those for the constitutive promoter (see Eqs. (13) and (17)), and the influence the two-state promoter is characterized by B_R and B_G . The bigger the translational burst size, the smaller the noise of the proteins, which is the same as the operon with the constitutive promoter.

3.2.3. The correlation between mRNAs and the correlation between proteins

For the operon with the constitutive promoter, both synthesis and degradation of RG and R are independent with each other, so numbers of RG and R are uncorrelated. However, for the operon with the two-state promoter, mRNAs RG and R are synthesized only when the promoter is in the on state, and both mRNAs are just under degradation in the off state. This will bring correlation between numbers of RG and R, which is clearly shown in the stochastic simulation (Fig. 4B). The correlation between RG and R

depends on the dynamics parameters of the two-state promoter: k_{on} , k_{off} and k_{TX} . We will get the analytical result for the correlation and discuss the influences of these parameters.

When $\gamma_{RG} = \gamma_R$, then $A_R = A_{RG} = A$, we will get

$$\rho_{RG-R} = \frac{1}{\sqrt{\left(1 + \frac{1}{Ae}\right)\left(1 + \frac{1}{A(1-e)}\right)}}, \quad (45)$$

$$\rho_{mR-mG} = \frac{\sqrt{1-e}}{\sqrt{1 - \frac{e}{1 + \frac{1}{A}}}}, \quad (46)$$

where

$$A = \frac{k_{off}}{k_{on} + k_{off}} \frac{k_{TX}}{(\gamma_{RG} + k_{on} + k_{off})}.$$

Unlike the operon with the constitutive promoter where $\rho_{RG-R} = 0$, mRNAs RG and R transcribed from a two-state promoter have nontrivial correlation which is dependent on e and A . If we adjust the dynamics parameters of the two-state promoters, the correlation will increase as A become larger. The first product term of A is $f_{off} = k_{off}/(k_{on} + k_{off})$ which is the possibility that promoter is in the off state. The effects of f_{off} on the correlation are shown in Fig. 4A and B. If the promoter

stays in the on state too long, the correlation between RG and R will be disturbed by the independent synthesis of RG and R (Fig. 4A). The second product term is $k_{TX}/(\gamma_{RG} + k_{on} + k_{off})$. It quantifies the influence of the transcriptional burst size (k_{TX}/k_{off}) on the correlation. If the transcriptional burst size is very small, the possibility that both RG and R are synthesized in a period of on state will be small. This leads to small correlation between the numbers of RG and R (Fig. 4C, both the burst sizes of RG and R are 0.15). Taken together all of these results, we get the conclusion that the more time promoter stays in the off state and the bigger transcriptional burst size, the bigger the correlation between mRNAs. When $A \gg 1$, $\rho_{mR-mG} \rightarrow 1$, which means promoter dynamics strongly influences the correlation.

When $\gamma_{RG} = \gamma_R$ and $\gamma_{PR} = \gamma_{PG}$, then $B_R = B_G = B$, the correlation between proteins is

$$\begin{aligned} \rho_{RFP-GFP} &= \frac{\sqrt{1-e}}{\sqrt{\left[1 + \frac{1}{k_{PR}}(\gamma_{RG} + \gamma_{PG})\right]\left[1 + \frac{1}{k_{PG}}(\gamma_{RG} + \gamma_{PG})\right] - \frac{e}{1+\frac{1}{B}}}} \\ &\approx \frac{\sqrt{1-e}}{\sqrt{\left(1 + \frac{1}{b_{PR}(1+B)}\right)\left(1 + \frac{1}{b_{PG}(1+B)} - \frac{e}{1+\frac{1}{B}}\right)}}} \end{aligned} \quad (47)$$

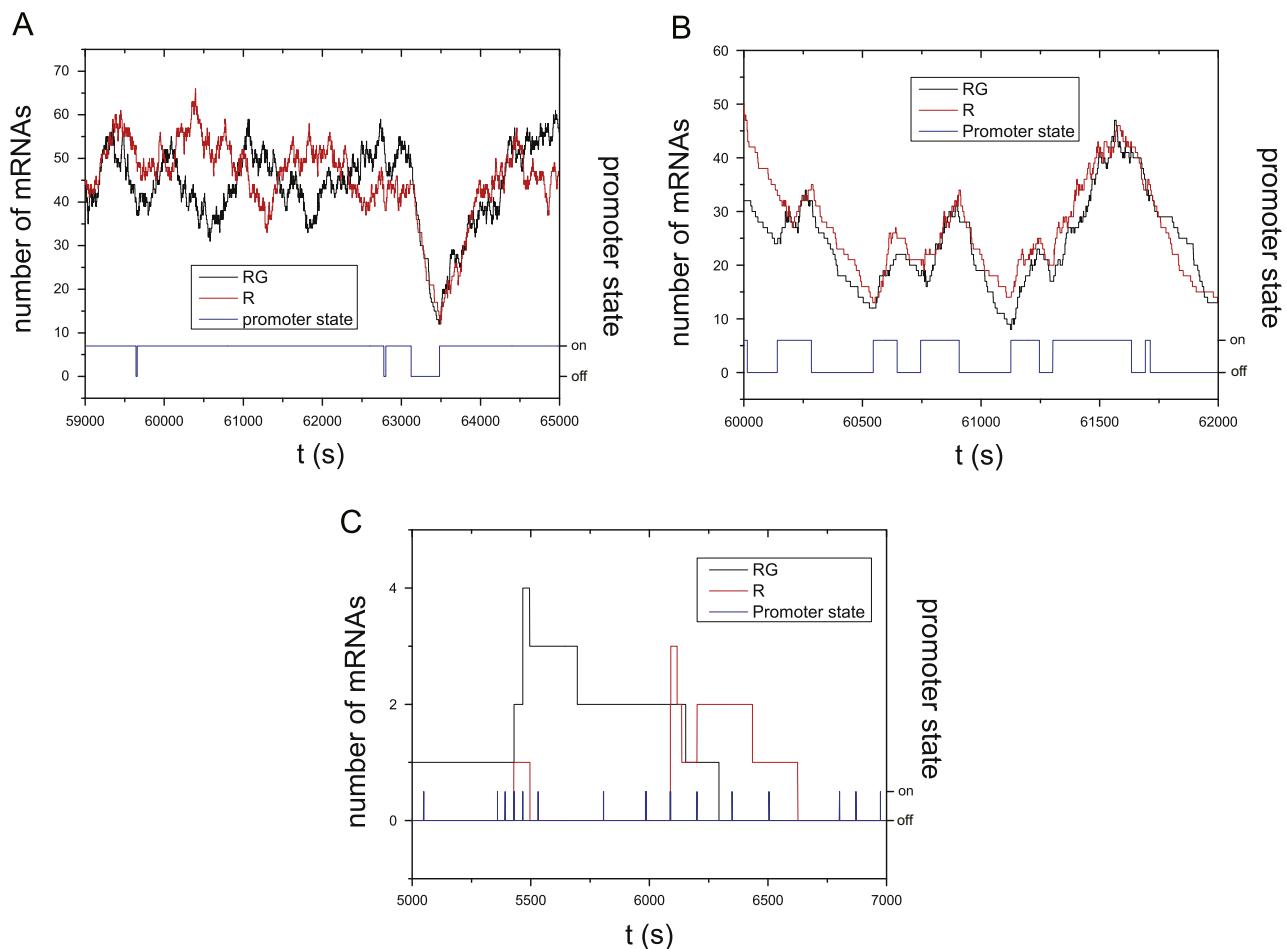


Fig. 4. Sampling of gene expression profiles from stochastic simulation for the operon with the two-state promoter. Promoter stays in the on state for long time (A) or small transcription burst size (C) would lower the correlation between two mRNAs. In the simulations $\gamma_R = \gamma_{RG} = 0.0033 \text{ s}^{-1}$, $e = 0.5$, $k_{on} = 0.006 \text{ s}^{-1}$, $k_{TX} = 0.3 \text{ s}^{-1}$; other parameters in (A) $k_{off} = 0.00075 \text{ s}^{-1}$, $f_{off} \approx 0.11$, the average transcriptional burst sizes of RG and R are $b_{RG} = b_R = 200$; (B) $k_{off} = 0.006 \text{ s}^{-1}$, $f_{off} = 0.5$, $b_{RG} = b_R = 25$; (C) $k_{off} = 1 \text{ s}^{-1}$, $f_{off} \approx 1$, $b_{RG} = b_R = 0.15$.

When $\gamma_{RG} \gg \gamma_{PG}$,

$$B \approx \frac{k_{off}}{k_{on} + k_{off}} \frac{k_{TX}}{(\gamma_{PG} + k_{on} + k_{off})}.$$

We can see the correlation between proteins is bigger than in the operon with the constitutive promoter (Eq. (24)). Changing dynamics parameters of the two-state promoters could change the correlation between proteins. If we adjust the parameters of two-state promoter and let $B \rightarrow 0$, correlation between proteins will be the same for the constitutive promoter.

3.2.4. Effects of dynamics parameters on the correlation

Among three dynamics parameters of the two-state promoter, *E. coli* cell seems to adjust k_{off} to modulate the transcription strength (So et al., 2011). We first investigate how the correlation depends on k_{off} . Both the stochastic simulation and analytical results are shown in Fig. 5A and B. They agree with each other very well. As we discussed above, the two-state promoter would increase the correlation between proteins compared with the constitutive promoter. When k_{off} decreases to zero (Fig. 4A), the promoter dynamics will be the same as the constitutive promoter, and the promoter will not be turned off. When $k_{off} \gg k_{on}$ (Fig. 4C), the promoter stays in the on state only for a very short time and the transcriptional burst sizes are very small. So the possibility that both RG and R are synthesized in same period of on state will be small, and the correlation between RG and R will be lost. Between these two limits, the correlation would increase first and then decrease as k_{off} increase further (Fig. 5A and B). The exact theoretical k_{off} at which $\rho_{RFP-GFP}$ gets the maximal value is a little complicated for general case. But from Eq. (47), we can see that $\rho_{RFP-GFP}$ get its maximal value when B get the maximal value for the case of $\gamma_R = \gamma_{RG}$ and $\gamma_{PR} = \gamma_{PG}$. When $\gamma_{RG} \gg \gamma_{PG}$,

$$B \approx \frac{k_{off}}{k_{on} + k_{off}} \frac{k_{TX}}{(\gamma_{PG} + k_{on} + k_{off})}$$

will get the maximal value at $k_{off} = \sqrt{k_{on}(k_{on} + \gamma_{PG})}$, which means the correlation between RFP and GFP will get the maximal value at this point.

We also study the correlation between proteins with different dynamics parameters while keeping the average expression level constant. When we adjust k_{off} or k_{on} , we adjust k_{TX} accordingly to keep the effective transcription rate $K = k_{on}k_{TX}/(k_{on} + k_{off})$ unchanged. Analytical results and stochastic simulations for the correlation between proteins are shown in Fig. 5C–F. If we adjust the dynamics parameters of transcription, the correlation between the numbers of mRNAs coding for RFP and GFP will change, and the correlation between proteins varies accordingly. In the case of fixing k_{on} and increasing k_{off} , the duration time of one transcription burst decreases, and the time that the promoter stays in the off state are not changed. The average numbers of mRNAs R and RG are unchanged, but these mRNAs are synthesized in a series of shorter time windows and in other times are just degraded. This will lead to larger correlation between mRNAs, which will transmit to proteins level (Fig. 5C and D). Based on similar reason, fixing k_{off} and increasing k_{on} will decrease the correlation between the number of mRNAs coding for RFP and GFP, and then decrease the correlation between proteins (Fig. 5E and F).

In Fig. 5A, C and E, a variation of terminator efficiency does not significantly change the correlation between proteins in some regimes. This is because the correlation between R and RG is significant in these regimes (see Eq. (46)), so the ratio of RG in the total mRNA (determined by e) would not obviously influence the correlation between proteins. This phenomenon is one of the main differences between these two kinds of promoters. Outside of these regimes, the polar effect will affect the correlation between proteins obviously: the larger the termination efficiency, the

smaller the correlation (see Fig. 5A, C, and E). However, changing the translation rate in the normal range ($0.01\text{--}1\text{ s}^{-1}$) almost does not vary the correlation (see Fig. 5B, D, and F). The dependence is obvious only when the translation rate is very small (below 0.001 s^{-1}) which is quite rare for protein synthesis in bacteria. So adjusting translation rate is still a better strategy than adjusting termination efficiency to tune expression ratio of genes in the operon. This is the same with the constitutive promoter.

4. Discussion

In the post-genome era, understanding the relationship between gene sequences, genes organization and biological functions is a great challenge. In this work, we study noise of gene expression in the bicistronic operon with polar effect. We try to figure out how different parameters affect the correlation of these two gene expressions. Our result may be helpful to understand which mechanism cell chooses to coordinate gene expression levels and why.

Previous experiments have shown that the expression correlation between two genes in the same operon is stronger than those in two separate operons even with identical promoters (Ray and Igoshin, 2012; Tabor et al., 2008). This strengthened correlation comes from co-transcription of the two genes (Sneppen et al., 2010; Swain, 2004; Ray and Igoshin, 2012; Iber, 2006). It makes sense that intergenic terminator would lower the correlation. If co-translation is not considered, translation processes of different genes in the same operon are independent with each other. This may lead the correlation of gene expressions to be further reduced at the protein level.

For the large protein complex synthesis, subunits need to be produced in a determined ratio to avoid wasting and to increase the efficiency (Carmi et al., 2006, 2009; Swain, 2004). As shown in Figs. 3 and 5, if the cell tunes the subunit expression level in transcription level using terminator in the intergenic region, the correlation of these two proteins would be lowered, which is not favored by the cell. However, altering the translation rate almost does not alter the correlation in a large parameter space (Figs. 3 and 5). So tuning the protein levels with the translation rate would be a better strategy. Actually, recent work by Quax et al. (2013) shows that different translation rates are a key factor determining expression levels of genes in an operon. Correlative expressions of enzymes in the same metabolic pathway will avoid intermediate cumulation which may be harmful. Thus putting these enzyme genes in an operon without obvious polar effect would benefit cell growth (Ray and Igoshin, 2012).

Using bioinformatic methods, it was found that there are a lot of terminators in the intergenic region of the operons (Lesnik et al., 2001). It is practical that a cell may tune the expression ratio at the transcriptional level for those genes in an operon by inserting terminators between genes. So there is a balance between the advantage of co-regulation in the same operon and the detrimental effect of lowered correlation by the polar effect. For those circumstances that function are not sensitive with the correlation, operon with intergenic terminator would be one of the choices to tune the expression ratio. For example, *recA* and *recX* genes form an operon with an intergenic terminator in *E. coli* (Pagès et al., 2003). *RecA* and *RecX* are two proteins with the opposite functions in the process of repairing of double strand DNA breaking. Normal cell function needs both of them (Stohl et al., 2003). *RecA* forms poly-*RecA* to repair DSB (double strand breaking), and *RecX* can inhibit function of *RecA* as the ratio of *RecX* to *RecA* varies in a broad range (Stohl et al., 2003). So low correlation between them would have little impact on their function. It is no wonder why a

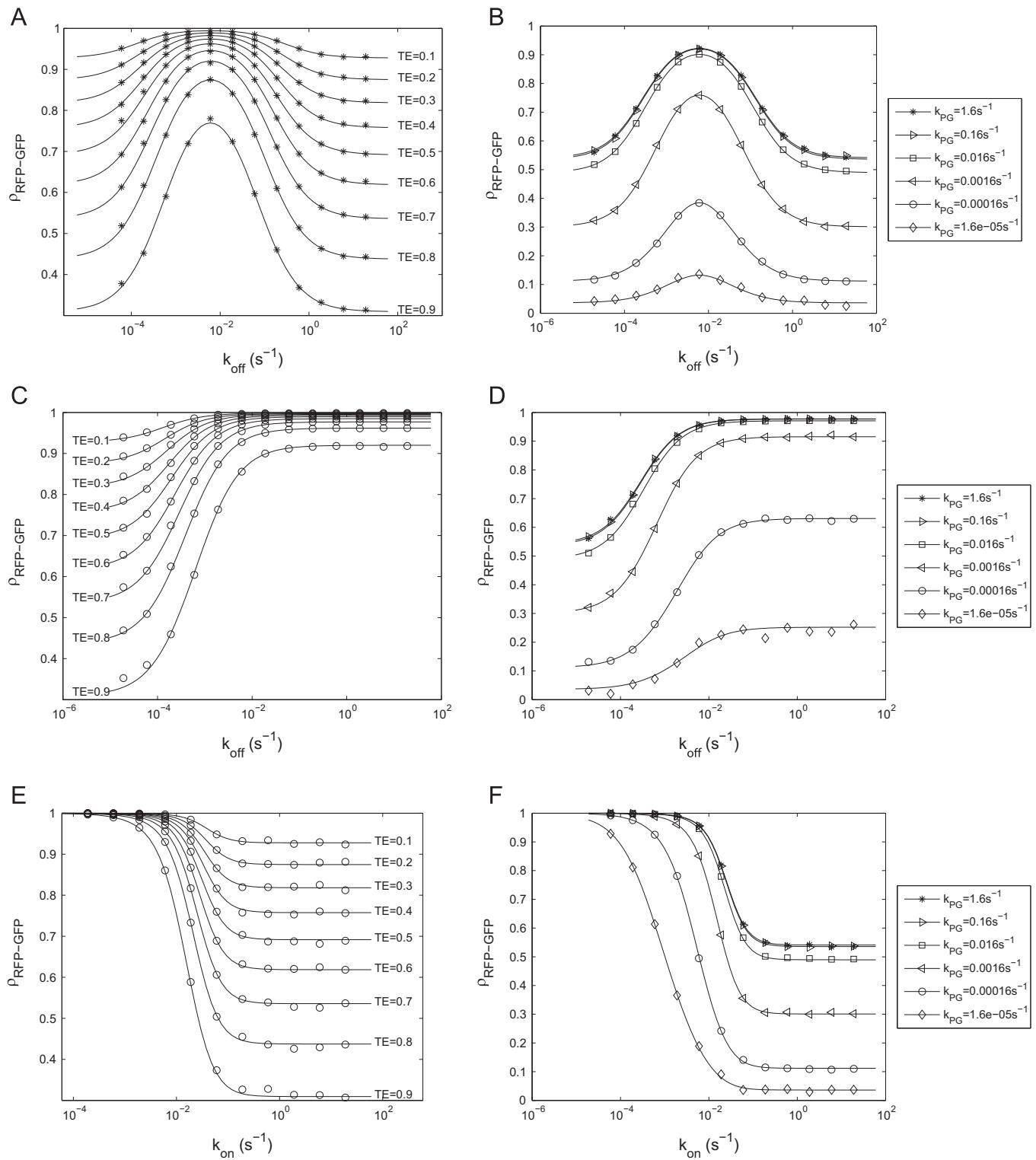


Fig. 5. Effects of polar effect and translation rate on the correlation between RFP and GFP expressed from a bicistronic operon driven by a two-state promoter. Lines are the analytical results and symbols are from stochastic simulations. (A,B)(C,D)(E,F) are correspondent with three ways of modulation of the two state promoter. In (A,B) k_{on} , k_{tx} are fixed, only k_{off} is changed; in (C,D) k_{on} and the average mRNA expression levels are fixed, k_{off} is changed and k_{tx} is changed correspondingly; in (E,F) k_{off} and the average mRNA expression levels are fixed, k_{on} is changed and k_{tx} is changed correspondingly. Parameters used are as follows: $\gamma_{PR} = \gamma_{PG} = 0.00029 \text{ s}^{-1}$, $\gamma_R = \gamma_{RG} = 0.0033 \text{ s}^{-1}$, $K_{PR} = 0.16 \text{ s}^{-1}$, in (A,C,E) $k_{PG} = 0.16 \text{ s}^{-1}$; (B,D,F) $e = 0.7$; (A,B) $k_{on} = 0.006 \text{ s}^{-1}$, $k_{tx} = 0.3 \text{ s}^{-1}$; (C,D) $k_{on} = 0.006 \text{ s}^{-1}$; (E,F) $k_{on} = 0.006 \text{ s}^{-1}$.

cell utilizing a terminator with $e \approx 0.9$ maintains a relatively lower level of RecX with RecA without affecting their functions.

Changing mRNA stability can tune the protein expression ratio, but it will also change the protein correlation significantly. Increasing lifetime of full-length mRNA will enhance the correlation, and

increasing lifetime of the short one gets the opposite effect. It is because a longer lifetime of full-length mRNA would lead to a bigger translation burst, and a larger ratio of protein produced from the full-length mRNA, both of which can increase the correlation. Increasing the lifetime of short mRNA just lowers the ratio

of uncorrelated RFP from total RFP protein. However, the cell seems to avoid tuning the mRNA stability for different parts of one operon when the proteins they code have related function (Bernstein et al., 2002; Kristoffersen et al., 2012; Quax et al., 2013).

Intrinsic and extrinsic noises together determine gene expression profiles (Taniguchi et al., 2010; Rosenfeld et al., 2005; Paulsson, 2004). In this work, we just consider the former. The intrinsic noise is the main source of noise when the average number of protein is low (Taniguchi et al., 2010; Elowitz et al., 2002). For a single cell, extrinsic noise varies relatively slowly. Experiments show it would introduce correlation for independent expressed proteins in the time of 20 min to one cell-cycle (Taniguchi et al., 2010; Rosenfeld et al., 2005). However, 20 min is a long time for fast growing cells, so our intrinsic results would not lose the sense.

Genes work as a network. So to understand the genome design principle we need to consider interactions between genes, including the influence from the stochasticity of gene expression. Our analytical results can give direct estimation for the relation of noise and correlation of gene expression, which may help to understand the relationship between genome structure and robust gene interaction network. Recent development of single cell or molecule experiment offers quantitative measurement of gene expression in time and space dimension (Taniguchi et al., 2010; So et al., 2011; Nevo-Dinur et al., 2011). Our results may help to design and understand this kind of experiments.

Acknowledgments

The work was supported by the National Basic Research Program of China (Grant no. 2013CB834100), National Natural Science Foundation of China (Grant nos. 11121403 and 11274320) and Open Project Program of State Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, China (no. Y4KF171CJ1). H.S. thanks Weimou Zheng for useful discussions. L.X. acknowledges support from National Natural Science Funds for Young Scholar (no. 11304115) and support from China Postdoctoral Science Foundation (no. 2013M541282).

Appendix A. Parameter estimation for gene expression in *E. coli*

1. *The mRNA degradation rate*: about 80% of mRNAs have a half-life time 3–8 min (Bernstein et al., 2002). We use $\beta_{mRNA} = 1/5 \text{ min}^{-1} \sim 0.0033 \text{ s}^{-1}$ by default, and vary mRNAs half-life just in the range of 0.5–15 min when we discuss the parameter dependence.
2. *The protein degradation rate*: protein is only under growth dilution in bacteria, almost no degradation, so $\beta_{protein} = \ln 2/T = 0.00029 \text{ s}^{-1}$, T is the doubling time of bacterial cell, we choose $T=40$ min.
3. *The translation rate*: considering the limit for translation elongation and the time needed for translation initiation (about 1 s, Proshkin et al., 2010), the fastest translation rate should be around 1 s^{-1} , the smallest we use 0.0011 s^{-1} (average one translation happens from the most stable mRNA during its 15 min half-life time). The average translation rate is in the range of $0.1\text{--}0.3 \text{ s}^{-1}$ (Bremer and Dennis, 1996). For *E. coli* with doubling time $T=40$ min, the average translation rate is around 0.16 s^{-1} which we used as default (Bremer and Dennis, 1996).
4. *The transcription rate*: the highest estimation 1 s^{-1} is from rrn genes transcription for the fast growing *E. coli* cell (Bremer and Dennis, 1996).

5. *The two-state promoter parameters*: from Fig. S12C and D of So et al. (2011), we use $k_{on} = 0.006 \text{ s}^{-1}$, $k_{off} = 0.006 \text{ s}^{-1}$, $k_{TX} = 0.3 \text{ s}^{-1}$ as default.

Appendix B. Stochastic simulation method

The Gillespie (1977) algorithm, developed by Daniel T. Gillespie in 1970s, is well known as an effective stochastic simulation algorithm. By calculating the reaction probability density function (RPDF) $P(\tau, \mu)$ using the current reaction rate and reactant species combination number, probability of when the next reaction will occur (τ) and which reaction will occur (μ) can be provided. Then we renew the molecular number of relevant species. In this way, we get a statistically correct evolution trajectory of the chemical system, which is rigorously equivalent to that produced by the distribution function $f(q_1, \dots, q_n; t)$ as the solution of the master equation. We use Gillespie algorithm (Gillespie, 1977; Hao et al., 2011a) to simulate the biochemical reactions shown in Fig. 1. Parameters used in the simulations are shown in Appendix A. We first generate a long enough stochastic evolution trajectory for each chemical species. With these trajectories, we start the sampling after the system reaches the steady state (the waiting time we used is 10 protein half-lives). In order to make sure that the statistic points are fully independent of each other, we select the points in the trajectories with the interval of 2 protein half-lives time, and then calculate the correlation between different species. Each data point of the simulation results is the result of at least 10,000 trials.

References

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., Blüthgen, N., 2013. Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* 9, 675.

Berg, O.G., 1978. A model for the statistical fluctuations of protein numbers in a microbial population. *J. Theor. Biol.* 71, 587–603.

Bernstein, J.A., Khodursky, A.B., Lin, P.-H., Lin-Chao, S., Cohen, S.N., 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. USA* 99, 9697–9702.

Bremer, H., Dennis, P.P., 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhardt, F.C. (Ed.), *Escherichia coli* and *Salmonella*. ASM Press, Washington, D.C., pp. 1553–1569.

Cannarozzi, G., Schraudolph, N.N., Faty, M., vonRohr, P., Friberg, M.T., Roth, A.C., Gonnet, P., Gonnet, G., Barral, Y., 2010. A role for codon order in translation dynamics. *Cell* 141, 355–367.

Carmi, S., Levanon, E., Havlin, S., Eisenberg, E., 2006. Connectivity and expression in protein networks: proteins in a complex are uniformly expressed. *Phys. Rev. E* 73, 031909.

Carmi, S., Levanon, E.Y., Eisenberg, E., 2009. Efficiency of complex production in changing environment. *BMC Syst. Biol.* 3, 3.

Chalancón, G., Ravarani, C.N.J., Balaji, S., Martinez-Arias, A., Aravind, L., Jothi, R., Babu, M.M., 2012. Interplay between gene expression noise and regulatory network architecture. *Trends Genet.* 28, 221–232.

Chen, H., Bjerknes, M., Kumar, R., Jay, E., 1994. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res.* 22, 4953–4957.

Cho, B.-K., Zengler, K., Qiu, Y., Park, Y.S., Knight, E.M., Barrett, C.L., Gao, Y., Palsson, B.Ø., 2009. The transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotechnol.* 27, 1043–1049.

Chong, S., Chen, C., Ge, H., Xie, X.S., 2014. Mechanism of transcriptional bursting in bacteria. *Cell* 158, 314–326.

Condon, C., 2007. Maturation and degradation of RNA in bacteria. *Curr. Opin. Microbiol.* 10, 271–278.

Dandekar, T., Snel, B., Huynen, M., Bork, P., 1998. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* 23, 324–328.

Eldar, A., Elowitz, M.B., 2010. Functional roles for noise in genetic circuits. *Nature* 467, 167–173.

Elowitz, M.B., Levine, A.J., Siggia, E.D., Swain, P.S., 2002. Stochastic gene expression in a single cell. *Science* 297, 1183–1186.

Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361.

Golding, I., Paulsson, J., Zawilski, S.M., Cox, E.C., 2005. Real-time kinetics of gene activity in individual bacteria. *Cell* 123, 1025–1036.

Goodman, D.B., Church, G.M., Kosuri, S., 2013. Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342, 475–479.

Güell, M., van Noort, V., Yus, E., Chen, W.-H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kühner, S., Rode, M., Suyama, M., Schmidt, S., Gavin, A.-C., Bork, P., Serrano, L., 2009. Transcriptome complexity in a genome-reduced bacterium. *Science* 326, 1268–1271.

Gur, E., Biran, D., Ron, E.Z., 2011. Regulated proteolysis in Gram-negative bacteria—how and when? *Nat. Rev. Microbiol.* 9, 839–848.

Hao, Y., Xu, L., Shi, H., 2011a. Theoretical analysis of catalytic-sRNA-mediated gene silencing. *J. Mol. Biol.* 406, 195–204.

Hao, Y., Zhang, Z.J., Erickson, D.W., Huang, M., Huang, Y., Li, J., Hwa, T., Shi, H., 2011b. Quantifying the sequence–function relation in gene silencing by bacterial small RNAs. *Proc. Natl. Acad. Sci. USA* 108, 12473–12478.

Hiroe, A., Tsuge, K., Nomura, C.T., Itaya, M., Tsuge, T., 2012. Rearrangement of gene order in the phaCAB operon leads to effective production of ultrahigh-molecular-weight poly[(R)-3-hydroxybutyrate] in genetically engineered *Escherichia coli*. *Appl. Environ. Microbiol.* 78, 3177–3184.

Iber, D., 2006. A quantitative study of the benefits of co-regulation using the *spolI* operon as an example. *Mol. Syst. Biol.* 2, 43.

Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3, 318–356.

Klumpp, S., Dong, J., Hwa, T., 2012. On ribosome load, codon bias and protein abundance. *PLoS One* 7, e48542.

Kristoffersen, S.M., Haase, C., Weil, M.R., Passalacqua, K.D., Niazi, F., Hutchison, S.K., Desany, B., Kolstø, A.-B., Tourasse, N.J., Read, T.D., Økstad, O.A., 2012. Global mRNA decay analysis at single nucleotide resolution reveals segmental and positional degradation patterns in a Gram-positive bacterium. *Genome Biol.* 13, R30.

Lesnik, E.A., Sampath, R., Levene, H.B., Henderson, T.J., McNeil, J.A., Ecker, D.J., 2001. Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Res.* 29, 3583–3594.

Liang, L.W., Hussein, R., Block, D.H.S., Lim, H.N., 2013. Minimal effect of gene clustering on expression in *Escherichia coli*. *Genetics* 193, 453–465.

Li, G.-W., Oh, E., Weissman, J.S., 2012. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* 484, 538–541.

Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S., Amster-Choder, O., 2011. Translation-independent localization of mRNA in *E. coli*. *Science* 331, 1081–1084.

Nishizaki, T., Tsuge, K., Itaya, M., Doi, N., Yanagawa, H., 2007. Metabolic engineering of carotenoid biosynthesis in *Escherichia coli* by ordered gene assembly in *Bacillus subtilis*. *Appl. Environ. Microbiol.* 73, 1355–1361.

Pagès, V., Koffel-schwartz, N., Fuchs, R.P.P., 2003. *recX*, a new SOS gene that is co-transcribed with the *recA* gene in *Escherichia coli*. *DNA Repair (Amst)* 2, 273–284.

Paulsson, J., 2004. Summing up the noise in gene networks. *Nature*, 415–418.

Paulsson, J., Berg, O.G., Ehrenberg, M., 2000. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proc. Natl. Acad. Sci. USA* 97, 7148–7153.

Peccoud, J., Ycart, B., 1995. Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.* 48, 222–234.

Price, M.N., Huang, K.H., Arkin, A.P., Alm, E.J., 2005. Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome Res.* 15, 809–819.

Proshkin, S., Rahmouni, A.R., Alexander, M., Nudler, E., 2010. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328, 504–508.

Quax, T.E.F., Wolf, Y.I., Koehorst, J.J., Wurtzel, O., van der Oost, R., Ran, W., Blombach, F., Makarova, K.S., Brouns, S.J.J., Forster, A.C., Wagner, E.G.H., Sorek, R., Koonin, E.V., van der Oost, J., 2013. Differential translation tunes uneven production of operon-encoded proteins. *Cell Rep.* 4, 938–944.

Ray, J.C.J., Igoshin, O.A., 2012. Interplay of gene expression noise and ultrasensitive dynamics affects bacterial operon organization. *PLoS Comput. Biol.* 8, e1002672.

Rosenfeld, N., Young, J., Alon, U., Swain, P., Elowitz, M., 2005. Gene regulation at the single-cell level. *Science* 307, 1962–1965.

Salgado, H., Moreno-Hagelsieb, G., 2000. Operons in *Escherichia coli*: genomic analyses and predictions. *Proc. Natl. Acad. Sci. USA* 97, 6652–6657.

Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M., Rosenow, C., 2003. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.* 13, 216–223.

Sneppen, K., Pedersen, S., Krishna, S., Dodd, I., Semsey, S., 2010. Economy of operon formation: cotranscription minimizes shortfall in protein complexes. *MBio* 1 e00177–10.

So, L.-H., Ghosh, A., Zong, C., Sepúlveda, L.A., Segev, R., Golding, I., 2011. General properties of transcriptional time series in *Escherichia coli*. *Nat. Genet.* 43, 554–560.

Stohl, E.A., Brockman, J.P., Burkle, K.L., Morimatsu, K., Kowalczykowski, S.C., Seifert, H.S., 2003. *Escherichia coli* *RecX* inhibits *RecA* recombinase and coprotease activities in vitro and in vivo. *J. Biol. Chem.* 278, 2278–2285.

Swain, P.S., 2004. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J. Mol. Biol.* 344, 965–976.

Swain, P.S., Elowitz, M.B., Siggia, E.D., 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA* 99, 12795–12800.

Tabor, J.J., Bayer, T.S., Simpson, Z.B., Levy, M., Ellington, A.D., 2008. Engineering stochasticity in gene expression. *Mol. Biosyst.* 4, 754–761.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., Xie, X.S., 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329, 533–538.

Thattai, M., van Oudenaarden, A., 2001. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA* 98, 8614–8619.

Tuller, T., Carmi, A., Vetsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., Pilpel, Y., 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141, 344–354.

Vellanoweth, R., Rabinowitz, J., 1992. The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo. *Mol. Microbiol.* 6, 1105–1114.

Wen, J.-D., Lancaster, L., Hodges, C., Zeri, A.-C., Yoshimura, S.H., Noller, H.F., Bustamante, C., Tinoco, I., 2008. Following translation by single ribosomes one codon at a time. *Nature* 452, 598–603.